

## DEMAND BASED GENERATION OF SYMBOLIC INFORMATION

### BACKGROUND OF THE INVENTION

#### 1. Field of the Invention

The present invention generally relates to computer aided software engineering (CASE) and, more particularly, to demand based generation of symbolic debugger information which provides an interactive and dynamic environment for computer program building and debugging. The invention allows a programmer to debug their programs without incurring the cost of generating symbolic information. Debugger symbolic information refers to the information that a program compiler communicates to the program debugger. This information describes to the debugger how to show the user a program's variable values, variable types, program counter, and stack backtrace. The costs of generating symbolic information include performance cost at compile time, storage cost associated with the symbolic information overhead, startup time for the debugger, variable value and program location access time while debugging, and linkage time associated with building the information associated with symbolic debugging. Demand based generation of symbolic information reduces the costs by eliminating additional compilation overhead until necessary, by removing additional storage requirements for symbolic debugging data, by matching and retrieving source to ensure correctness of debugging, and by enabling debugging at all times.

#### 2. Description of the Prior Art

Object oriented programming (OOP) is the preferred environment for building user-friendly, intelligent computer software. Key elements of OOP are data encapsulation, inheritance and polymorphism. These elements may be used to generate a graphical user interface (GUI), typically characterized by a windowing environment having icons, mouse cursors and menus. While these three key elements are common to OOP languages, most OOP languages implement the three key elements differently.

Examples of OOP languages are Smalltalk and C++. Smalltalk is actually more than a language; it might more accurately be characterized as a programming environment. Smalltalk was developed in the Learning Research Group at Xerox's Palo Alto Research Center (PARC) in the early 1970s. In Smalltalk, a message is sent to an object to evaluate the object itself. Messages perform a task similar to that of function calls in conventional programming languages. The programmer does not need to be concerned with the type of data; rather, the programmer need only be concerned with creating the right order of a message and using the right message. C++ was developed by Bjarne Stroustrup at the AT&T Bell Laboratories in 1983 as an extension of C. The key concept of C++ is class, which is a user-defined type. Classes provide object oriented programming features. C++ modules are compatible with C modules and can be linked freely so that existing C libraries may be used with C++ programs.

The complete process of running a computer program involves translation of the source code written by the programmer to machine executable form, referred to as object code, and then execution of the object code. The process of translation is performed by an interpreter or a compiler. In the case of an interpreter, the translation is made at the time the program is run, whereas in the case of a compiler, the translation is made and stored as object code prior to running the program. That is, in the usual compile

and execute system, the two phases of translation and execution are separate, the compilation being done only once. In an interpretive system, such as the Smalltalk interpreter, the two phases are performed in sequence. An interpreter is required for Smalltalk since the nature of that programming environment does not permit designation of specific registers or address space until an object is implemented.

A compiler comprises three parts; the lexical analyzer, the syntax analyzer, and the code generator. The input to the lexical analyzer is a sequence of characters representing a high-level language program. The lexical analyzer divides this sequence into a sequence of tokens that are input to the syntax analyzer. The syntax analyzer divides the tokens into instructions and, using a database of grammatical rules, determines whether or not each instruction is grammatically correct. If not, error messages are produced. If correct, the instruction is decomposed into a sequence of basic instructions that are transferred to the code generator to produce a low-level language. The code generator is itself typically divided into three parts; intermediate code generation, code optimization, and code generation. Basically, the code generator accepts the output from the syntax analyzer and generates the machine language code.

To aid in the development of software, incremental compilers have been developed in which the compiler generates code for a statement or a group of statements as received, independent of the code generated later for other statements, in a batch processing operation. The advantage of incremental compiling is that code may be compiled and tested for parts of a program as it is written, rather than requiring the debugging process to be postponed until the entire program has been written. However, even traditional incremental compilers must reprocess a complete module each time.

Optimizing compilers produce highly optimized object code which, in many cases, makes debugging at the source level more difficult than with a non-optimizing compiler. The problem lies in the fact that although a routine will be compiled to give the proper answer, the exact way it computes that answer may be significantly different from that described in the source code. Some things that the optimizing compiler may do include eliminating code or variables known not to affect the final result, moving invariant code out of loops, combining common code, reusing registers allocated to variables when the variable is no longer needed, etc. Thus, mapping from source to object code and vice versa can be difficult given some of these optimizations. Inspecting the values of variables can be difficult since the value of the variable may not always be available at any location within the routine. Modifying the values of variables in optimized code is especially difficult, if not impossible. Unless specifically declared as volatile, the compiler "remembers" values assigned to variables and may use the "known" value later in the code without rereading the variable. A change in that value could, therefore, produce erroneous program results.

Once a program has been compiled and linked, it is executed and then debugged. Because logical errors, also known as "bugs," are introduced by programmers, they will want to detect and understand the errors, using a program debugger. After correcting the errors and recompiling, they use the debugger to confirm that those errors have been eliminated. Other uses for the debugger include inspecting executing programs in order to understand their operation, monitoring memory usage, instrumenting and testing programs, verifying the correctness of program translation by the compiler, verifying the correctness of operation of